Math 21 – Pointers for Section 3.2

There are 3 measures of dispersion (or spread) discussed in this section: range, standard deviation, and variance. They are designed to tell you how "spread out" a set of data is. In general, the larger the result, the more spread out the data are.

Range

The range is a quick way to determine dispersion. It is equal to maximum – minimum. There are some problems with range as your only measure. It is definitely not resistant to outliers. If you have an outlier in your data set, it will be the maximum or the minimum (or both). This could cause the set of data to seem to be more spread out than it really is.

For example, the set {1,2,3,4,4,5,6,7,200} has the same range as the set {1,20,45,62,87,109,125,146,167,200}, but they are definitely spread out in different ways.

Start with the range, but always look further.

Standard Deviation

This is the most used measure of dispersion in this class. Its calculation compares each value to the mean, so it really measures spread from the center, instead of from end to end as range does.

It would be great if we could say that the standard deviation has an interpretation such as the average distance of all values from the mean, but it does not. The thing to remember is that the larger the result, the more spread out the data are.

Rounding: By hand, always round the standard deviation to 2 decimal places beyond the accuracy of the numbers you have. In MyMathLab, pay attention to their directions when it comes to rounding.

If you are finding the standard deviation of a population, use the Greek symbol σ. (sigma)
If you are finding the standard deviation of a sample, use the symbol s.

You will always be able to use StatCrunch to calculate standard deviation, but be sure to look over the topic video for how it is calculated by hand. It will give you greater intuition for what it actually represents.

Sample or Population?

The two formulas are slightly different and will therefore give different answers, so you must know whether your set of data represents a sample or population.

Remember that the data are a sample if we only have part of the information we are interested in.

The data are a population if we have all of the information we are interested in.

Variance

Variance is simply equal to the standard deviation squared.

$$variance = (standard\ deviation)^2$$

If you are finding the variance of a population, use the symbol σ^2. (sigma squared)
If you are finding the variance of a sample, use the symbol s^2. (s squared)

StatCrunch

You can calculate these measures at the same time, just like mean and median.

- Enter all of the data in one column, 1 value per line.
- Press the Stat button, and select Summary Stats > Columns
- Select the column containing your data.
- Statistics, select the following.
  Range: Range (Highlighted by default)
  Sample Variance: Variance(Highlighted by default)
  Sample Standard Deviation: Std. dev. (Highlighted by default)
  Population Variance: Unadj. Variance
  Population Standard Deviation: Unadj. Std. dev.
- Click on Compute!

Empirical Rule

This rule uses standard deviation to tell us what percent of the values we should expect to find in a given interval. In this example I will talk about IQ scores, which have a mean of 100 points and a standard deviation of 15 points.

We expect to find 68% of the data within 1 standard deviation of the mean, from 15 points below the mean to 15 points above the mean.
So, we should expect 68% of IQ scores to be between 85 & 115.
This rule is symmetrical, so half of those values should be below the mean (34% between 85 & 100) while the other half should be above the mean (between 100 & 115).

We expect to find 95% of the data within 2 standard deviations (or 30 points in this example) of the mean. In other words, we expect 95% of the data to be between 70 & 130.

We expect 99.7% to lie within 3 standard deviations of the mean, which in this example would be between 55 & 145.

That leaves 0.3% outside of that range. Splitting it in half, that puts 0.15% below 55 and 0.15% above 145.

I recommend that you mark up the bell curve like I do in the Empirical Rule topic video.